(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平6－215049

(43)公開日　平成６年(1994)８月５日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　15/401		7218－5Ｌ		
15/20	５５０　Ｆ	9288－5Ｌ		

審査請求　未請求　請求項の数２　ＯＬ　（全７頁）

(21)出願番号　特願平5－7427

(22)出願日　平成５年(1993)１月20日

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町22番22号

(72)発明者　乾　隆夫
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　芥子　育雄
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　石鞍　謙一郎
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

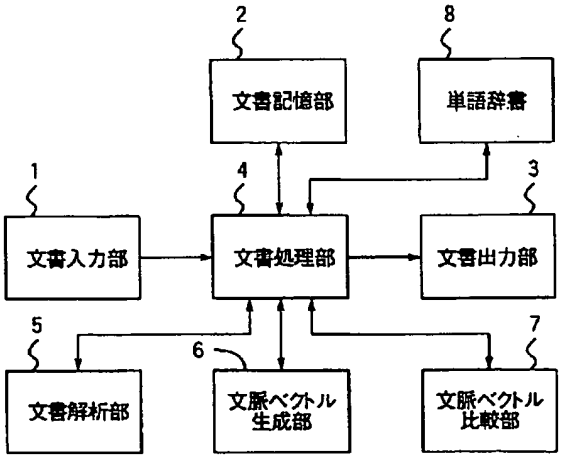
(74)代理人　弁理士　青山　葆　（外１名）

(54)【発明の名称】　文書要約装置

(57)【要約】

【目的】　特定の文書形式や文脈を仮定することなく簡単な処理によって文書における質の良い重要部分を抽出する。

【構成】　文書解析部５は文書入力部１から入力された文書を段落，文および単語に分解する。文脈ベクトル生成部６は、単語辞書８を用いて文，段落および文書の文脈ベクトルを生成する。文脈ベクトル比較部７は、文書と各段落毎の文，文書と各段落，段落と段落内の各文および文書と各文の文脈ベクトルを比較して各文脈ベクトル間距離算出する。文書処理部４は、各文脈ベクトル間距離を参照して、文書に最も近い段落と文書に近い複数文との２種類の要旨及び文書に最も近い各段落毎の文と各段落に最も近い文との２種類の要約を生成する。このように、入力文書を文脈ベクトルを用いて解析することによって、特定の文書形式や文脈を仮定することなく質の良い重要部分を簡単な処理で抽出できる。



1

【特許請求の範囲】

【請求項１】　単語の特徴ベクトルが格納された単語辞書と、

文書入力部から入力された文書に対して所定の解析を行って、上記入力文書を段落,文および単語に分割する文書解析部と、

上記分割された単語の特徴ベクトルを上記単語辞書を用いて求め、さらにこの求められた単語の特徴ベクトルに基づいて、上記分割された文および段落と上記入力文書の特徴ベクトルを所定の手順によって生成する特徴ベクトル生成部と、上記入力文書,段落および文の特徴ベクトル間の距離を所定の手順によって算出する距離算出部と、

上記算出された各特徴ベクトル間の距離に基づいて、上記入力文書の要約を所定の手順によって生成する文書要約生成部を備えたことを特徴とする文書要約装置。

【請求項２】　請求項１に記載の文書要約装置であって、

上記距離算出部は、上記入力文書と各段落との特徴ベクトル間距離,上記入力文書と各段落毎の文との特徴ベクトル間距離,各段落と夫々の段落内の文との特徴ベクトル間距離または上記入力文書と各文との特徴ベクトル間距離を算出し、上記文書要約生成部は、上記算出された各特徴ベクトル間距離に基づいて、上記入力文書に最も近い段落,上記入力文書に最も近い各段落毎の文,各段落に最も近い夫々の段落内の文および上記入力文書に近い複数の文の少なくとも一つを入力文書の要約として選出することによって上記入力文書の要約を生成することを特徴とする文書要約装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は、アイデアプロセッサや文書作成支援装置やワードプロセッサ等に使用されて文書作成や発想の支援を行う文書要約装置に関する。

【０００２】

【従来の技術】発想とは既知の情報の新たな組み合わせであり、決して無から有を作り出すことはできない。そのために、文書作成時における発想に際しては、既存の文書を参照して引用することが頻繁に行われる。

【０００３】一般に、参考とする既存の文書はその数も多く、個々の文書中における文章量も多い。したがって、この参考とする既存の文書をそのまま全部読んでいては時間や労力を消費してしまい、本来の目的である文書作成にかける力が減少してしまう。

【０００４】参考とする文書の多さについては、検索装置を用いて文書内容を絞り込むことによって減らすことができる。また、個々の文書中における文章量の多さについては、要約/要旨抽出装置を用いることによって減少できる。

【０００５】ここで、個々の文書の文章量を減少させることによって参照の手間を軽減するために、文書から要約/要旨抽出を抽出する場合を考える。この場合には、文書の文章量を減少させても元の文書に含まれる重要な内容が損なわれないような手法を用いる必要がある。

2

【０００６】従来から提唱されている文書要約の手法としては、次の２つの手法がある。第１の手法は、文章を表層的に解析するものである。この手法には、単語の出現頻度解析から文章の重要箇所を決定して元の文書に含まれている単語の組み合わせや文の抽出によって要約文の生成を行うものや、文の文末表現および用言によって文章中における強調/主張文を抽出するものが含まれる。

【０００７】第２の手法は、文章を意味的に解析するものである。この手法には、事前に文章の形式や文脈を仮定しておいてその仮定に沿って文章を解析して要約を抽出するものや、文の係り受けの粗密性を用いることによって内容の重要性を定義して要約を抽出するものが含まれる。

【０００８】

【発明が解決しようとする課題】上述のように、従来の文書要約の手法には、文章を表層的に解析する第１の手法と文章を意味的に解析する第２の手法との２つの手法があり、各手法には夫々以下のような問題点がある。すなわち、第１の手法の場合は、第２の手法に比べて簡単に実施できる反面、意味を扱わないので文書中の不要な箇所を重要な箇所と誤って判断してしまうという問題がある。一方、第２の手法の場合は、最初の仮定が当て嵌まらないようなタイプの文書に対しては全く非力であり、内容の重要性の定義自体が困難であるという問題がある。しかも、第１の手法に比べて処理が複雑である。

【０００９】そこで、この発明の目的は、特定の文書形式や文脈を仮定することなく、簡単な処理によって文書における質の良い重要部分を要約として抽出できる文書要約装置を提供することにある。

【００１０】

【課題を解決するための手段】上記目的を達成するため、第１の発明の文書要約装置は、単語の特徴ベクトルが格納された単語辞書と、文書入力部から入力された文書に対して所定の解析を行って上記入力文書を段落,文および単語に分割する文書解析部と、上記分割された単語の特徴ベクトルを上記単語辞書を用いて求め、さらにこの求められた単語の特徴ベクトルに基づいて上記分割された文および段落と上記入力文書の特徴ベクトルを所定の手順によって生成する特徴ベクトル生成部と、上記入力文書,段落および文の特徴ベクトル間の距離を所定の手順によって算出する距離算出部と、上記算出された各特徴ベクトル間の距離に基づいて、上記入力文書の要約を所定の手順によって生成する文書要約生成部を備えたことを特徴としている。

【００１１】また、第２の発明は、上記第１の発明の文

3

書要約装置であって、上記距離算出部は、上記入力文書と各段落との特徴ベクトル間距離,上記入力文書と各段落毎の文との特徴ベクトル間距離,各段落と夫々の段落内の文との特徴ベクトル間距離または上記入力文書と各文との特徴ベクトル間距離を算出し、上記文書要約生成部は、上記算出された各特徴ベクトル間距離に基づいて、上記入力文書に最も近い段落,上記入力文書に最も近い各段落毎の文,各段落に最も近い夫々の段落内の文および上記入力文書に近い複数の文の少なくとも一つを入力文書の要約として選出することによって上記入力文書の要約を生成することを特徴としている。

【００１２】

【作用】第１の発明では、文書入力部から文書が入力されると、文書解析部によって上記入力文書に対して例えば形態素解析等の解析が行われて上記入力文書が段落,文および単語に分割される。そして、この分割された単語の特徴ベクトルが特徴ベクトル生成部によって単語辞書を用いて求められ、さらにこの求められた単語の特徴ベクトルに基づいて、上記分割された文および段落と上記入力文書の特徴ベクトルが所定の手順によって生成される。

【００１３】そうすると、距離算出部によって、上記入力文書,段落および文の特徴ベクトル間の距離が所定の手順によって算出される。そして、この各特徴ベクトル間の距離に基づいて、文書要約生成部によって、上記入力文書の要約が所定の手順によって生成される。こうして、入力文書が特徴ベクトルを用いて解析されて文書における質の良い重要部が要約として抽出される。

【００１４】また、第２の発明では、特徴ベクトル生成部によって生成された入力文書,段落および文の特徴ベクトルに基づいて、距離算出部によって、上記入力文書と各段落との特徴ベクトル間距離,上記入力文書と各段落毎の文との特徴ベクトル間距離,各段落と夫々の段落内の文との特徴ベクトル間距離あるいは上記入力文書と各文との特徴ベクトル間距離が算出される。そして、この算出された各特徴ベクトル間距離に基づいて、文書要約生成部によって、上記入力文書に最も近い段落,上記入力文書に最も近い各段落毎の文,各段落に最も近い夫々の段落内の文および上記入力文書に近い複数の文の少なくとも一つが入力文書の要約として選出されて上記入力文書の要約が生成される。

【００１５】

【実施例】以下、この発明を図示の実施例により詳細に説明する。図１は本実施例の文書要約装置におけるブロック図である。文書入力部１は対話型のキーボードや光学式文字読み取り装置(ＯＣＲ)の他に通信回線や着脱式外部記憶装置で構成され、要約作成の対象となる文書が入力される。そして、文書入力部１から入力された文書は文書記憶部２に記憶される。さらに、この文書記憶部２には生成された要約文も格納される。

4

【００１６】文書出力部３は対話型のＣＲＴ(カソード・レイ・チューブ)や液晶表示装置(ＬＣＤ)の他にプリンタや通信回線や着脱式外部記憶装置で構成され、入力文書や要約文を出力する。

【００１７】文書処理部４は編集/検索等の一般的な文書処理を実施する他に、以下に述べる文書解析部５,文脈ベクトル生成部６及び文脈ベクトル比較部７を制御して、入力文書の要旨や要約を生成する。

【００１８】上記文書解析部５は、文書記憶部２から読み出した文書を解析して単語,文および段落に分解する。その際における文書解析方法としては、形態素解析を行って文書を単語に分解したり、特定の区切りに注目して文書を段落や文に分解したりする。上記文脈ベクトル生成部６は、上記文書解析部５によって文書を分解して得られた段落,文,単語および元の文書の文脈ベクトルを生成する。

【００１９】ここで、上記文脈ベクトルについて簡単に説明する。何個かの特徴語を用意して特徴空間を定義する。上記特徴語としては、例えば次のような単語を定義する。人間，男，女，機械，知識，活動，経験，政治，芸術，科学，…上記特徴語の個数は任意であるが、少なくとも２００語～５００語程度は用意しておく方が実用上は望ましい。また、特徴語の種類や分野についても任意であり、選択に当たっての厳密さは要求されず、特徴が相互にオーバーラップしていても構わない。さらに、要約抽出の対象となる文書の分野が特定の分野である場合には、その分野に特有の特徴語を充実させることによって、この特徴語に基づいて生成される文脈ベクトルの精度が向上して品質の高い要約を抽出できることになる。

【００２０】単語辞書８に文脈ベクトルを生成する際に使用される単語を格納し、上記単語辞書８に格納された各単語と上記特徴語との関連の有無(あるいは、関連の強度)に応じて当該単語を上記特徴空間に配置する。その際における各単語の特徴空間上の位置がその単語の文脈ベクトルであり、この文脈ベクトルは単語に対応付けて単語辞書８に格納される。

【００２１】図２は各単語の文脈ベクトルが定義された単語辞書８の内容の一例である。上記単語の文脈ベクトルは、単語辞書８内に格納されている単語と上記各特徴語との関連をその有無(あるいは、強度)によって段階的に表現した数字を要素とするベクトルである。すなわち、図２においては、関連がある場合には要素“１”を与え、関連が無い場合には要素“０”を与えている。尚、各要素の配列順序は上述した特徴語の配列順序と同じである。

【００２２】したがって、図２に例示された単語の文脈ベクトルは以下のことを表現している。すなわち、「人間」という単語は、各特徴語“人間”，“男”，“女”,…とは関連があり、各特徴語“機械”，“知識”，“活動”，“経

験″，“政治″，“芸術″，“科学″，…とは関連が無いと言う特徴を表現してる。また、「自動車」という単語は、各特徴語“人間″，“男″，“女″，“知識″，“経験″，“政治″，“芸術″，“科学″，…とは関連が無く、各特徴語“機械″，“活動″，…とは関連があると言う特徴を表現しているのである。

【００２３】尚、本実施例において文脈ベクトルを生成する際に用いる単語は、“名詞″および“サ変名詞(語尾に「する」と付けるとサ行変格活用動詞になる名詞)″だけである。したがって、単語辞書８に登録されている単語も名詞およびサ変名詞である。

【００２４】上記文脈ベクトル生成部６は、上記文書，段落，文および単語の文脈ベクトルを生成する際には次のようにして生成する。すなわち、先ず、上述のようにして予め単語辞書８に格納されている単語の文脈ベクトルを参照して、目的とする単語の文脈ベクトルを求める。次に、上述のようにして求められた目的とする文を構成する各単語(名詞およびサ変名詞)の文脈ベクトルを加算/正規化して、上記目的とする文の文脈ベクトルを求める。尚、上記文脈ベクトルの正規化とは、文脈ベクトルの長さを一定の値に揃えることである。

【００２５】また、上記段落の文脈ベクトルは、上述のようにして求められた目的とする段落を構成する各単語(名詞およびサ変名詞)の文脈ベクトルを加算/正規化して求める。同様に、目的とする文書を構成する各単語の文脈ベクトルを加算/正規化して、文書全体の文脈ベクトルを求める。

【００２６】上記文脈ベクトル比較部７は、上記文脈ベクトル生成部６によって生成された文書と各段落との文脈ベクトル，文書と各段落毎の文との文脈ベクトル，各段落と夫々の段落内の文との文脈ベクトルおよび文書と各文との文脈ベクトルの比較を行って、各文脈ベクトル間の距離を算出する。その際に、算出される２つの文脈ベクトル間の距離としては、正規化された当該両文脈ベクトルの内積を与える。そして、内積値が大きいほど距離が遠いとするのである。

【００２７】そして、上述のようにして算出された各文脈ベクトル間距離の値に基づいて、上記文書処理部４によって文書に近い段落や文および段落に近い文を選出することによって、入力文書の要約が生成されるのである。こうして生成された文書の要約は上記文書記憶部２に格納され、必要に応じて文書出力部３より出力される。

【００２８】すなわち、上記特徴ベクトルは文脈ベクトルであって、上記特徴ベクトル生成部を文脈ベクトル生成部６で構成し、上記距離算出部を文脈ベクトル比較部７で構成し、上記文書要約生成部を文書処理部４で構成するのである。

【００２９】図３は上記文書処理部４によって実施される要約作成処理動作のフローチャートである。以下、図３に従って上記要約作成処理動作について詳細に説明する。

【００３０】ステップＳ1で、上記文書入力部１から要約抽出の対象となる文書が入力されて文書記憶部２に記憶される。ステップＳ2で、上記文書解析部５によって、文書記憶部２から文書が読み出されて段落単位に分割される。その際に、例えば改行を段落の区切りとする。ステップＳ3で、上記文書解析部５によって、文書記憶部２から文書が読み出されて文単位に分割される。その際に、例えば句点を文の区切りとする。

【００３１】ステップＳ4で、上記文書解析部５によって、文書記憶部２から文書が読み出され、この読み出された文書が形態素解析によって単語に分解される。そして、得られた単語のうち名詞およびサ変名詞(以下、両者を単に単語という)のみが文書，上記ステップＳ2において分割された各段落および上記ステップＳ3において分割された各文の単位で文脈ベクトル生成部６に送出される。

【００３２】ステップＳ5で、上記文脈ベクトル生成部６によって、文書の文脈ベクトル，各段落の文脈ベクトルおよび各文の文脈ベクトルが次のようにして生成される。すなわち、先ず、上記文書を構成する単語，各段落を構成する単語および各文を構成する単語の文脈ベクトルが上記単語辞書８を引くことによって得られる。次に、各文を構成する単語の文脈ベクトルが加算され正規化されて各文の文脈ベクトルが得られる。同様に、各段落を構成する単語の文脈ベクトルが加算され正規化されて各段落の文脈ベクトルが得られ、文書を構成する単語の文脈ベクトルが加算され正規化されて文書の文脈ベクトルが得られる。

【００３３】ステップＳ6で、上記ステップＳ5において得られた文書の文脈ベクトルと各段落の文脈ベクトルとが比較されて各文脈ベクトル間の距離が算出される。また、文書の文脈ベクトルと各文の文脈ベクトルが各段落毎に比較されて各文脈ベクトル間の距離が算出される。また、各段落の文脈ベクトルと夫々の段落内の文の文脈ベクトルとが比較されて各文脈ベクトル間の距離が算出される。さらに、文書の文脈ベクトルと各文の文脈ベクトルとが比較されて各文脈ベクトル間の距離が算出される。ステップＳ7で、上記ステップＳ6において算出された文書と各段落との文脈ベクトル間距離が参照され、文書の文脈ベクトルに最も近い文脈ベクトルを有する段落が重要段落と見なされて、この重要段落が入力文書の要旨として文書記憶部２に格納され、必要に応じ文書出力部３から出力される。

【００３４】ステップＳ8で、上記ステップＳ6において算出された文書と各段落毎の文との文脈ベクトル間距離が参照され、文書の文脈ベクトルに最も近い文脈ベクトルを有する各段落毎の文が選出される。そして、選出された各段落毎の文が元の段落の順番に並べられて入力文

書の要約として文書記憶部２に格納され、必要に応じて文書出力部３から出力される。ステップＳ9で、上記ステップＳ6において算出された各段落と夫々の段落内の文との文脈ベクトル間距離が参照され、各段落の文脈ベクトルに最も近い文脈ベクトルを有する夫々の段落内の文が選出される。そして、上記選出された各段落毎の文が元の段落の順番に並べられて入力文書の要約として文書記憶部２に格納され、必要に応じて文書出力部３から出力される。ステップＳ10で、上記ステップＳ6において算出された文書と各文との文脈ベクトル間距離が参照され、文書の文脈ベクトルに最も近い文脈ベクトルを有する文から距離の短い順に所定数の文が選出される。そして、こうして選出された複数文が入力文書の要旨として文書記憶部２に格納され、必要に応じて文書出力部３から出力されて要約作成処理動作を終了する。

【００３５】オペレータは、上記文書出力部３から出力される２種類の要旨と２種類の要約から自分の目的に応じたものを選択して、以後の文書作成等に利用する。

【００３６】尚、上記文書出力部３は、通常の文書出力手段と同じに構成されている。したがって、上述の要旨/要約のみを出力したり、上述の要旨/要約の箇所がアンダーラインや反転等によって強調された文書全体を出力することが可能である。

【００３７】このように、上記実施例においては、文書入力部１から入力された文書を文書解析部５によって段落,文および単語に分解する。そして、文脈ベクトル生成部６によって上記文書を構成する単語,各段落を構成する単語および各文を構成する単語の文脈ベクトルを求め、この各単語の文脈ベクトルに基づいて各文の文脈ベクトル,各段落の文脈ベクトルおよび文書の文脈ベクトルを得る。そうした後、上記文脈ベクトル比較部７によって、各段落と文書との文脈ベクトル間距離,各段落毎の文と文書との文脈ベクトル間距離,各段落内の文と夫々の段落との文脈ベクトル間距離および各文と文書との文脈ベクトル間距離を算出する。

【００３８】そして、上記文書処理部４によって、文書に最も近い段落と文書に近い所定数の文との２種類の要旨、及び、文書に最も近い各段落毎の文の段落順の羅列と各段落に最も近い夫々の段落内の文の段落順の羅列との２種類の要約を生成して、上記文書出力部３から出力する。

【００３９】こうして、入力文書を文脈ベクトルを用いて解析することによって、従来の意味的解析を伴わない表層的な解析による上記第１の文書要約手法に比較して、文書における質の良い重要部分を抽出できる。また、従来の文章を意味的に解析する第２の文書要約手法に比較して、事前に特定の文書形式や文脈を仮定する必要がないので、入力文書に対する自由度が大きく種々のタイプの文書に適用可能である。さらに、入力文書の構造解析や文脈の意味理解を行って内容の重要性を定義する必要がないので、より簡単な処理によって要約の抽出を実施できる。

【００４０】上記実施例においては、各段落の文脈ベクトルは目的とする段落を構成する各単語の文脈ベクトルに基づいて求め、文書の文脈ベクトルはこの文書を構成する各単語の文脈ベクトルに基づいて求めている。しかしながら、この発明はこれに限定されるものではなく、各段落の文脈ベクトルは目的とする段落を構成する各文の文脈ベクトルに基づいて求め、文書の文脈ベクトルはこの文書を構成する段落の文脈ベクトルに基づいて求めてもよい。

【００４１】上記実施例においては、上記文脈ベクトル生成部６によって文脈ベクトルを生成する際に用いる単語辞書８に登録されている単語は名詞およびサ変名詞に限定しているが、この発明はこれに限定されないことは言うまでもない。また、上記実施例においては、文脈ベクトルの要素として当該単語と各特徴語とに関連がある場合には“１”を与える一方、関連が無い場合には“０”を与えている。しかしながら、この発明はこれに限定されるものではなく、関連の強度を段階的に表現した数字を与えてもよい。また、上記実施例においては、文書に最も近い段落,文書に近い所定数の文,文書に最も近い各段落毎の文の段落順の羅列および各段落に最も近い夫々の段落内の文の段落順の羅列から成る４種類の要旨/要約を生成して上記文書出力部３から出力するようにしているが、その中の幾つかを組み合わせて出力してもよい。

【００４２】上記実施例における文書要約装置は、必ずしも単独で使用しなければならない訳ではなく、従来からの文書要約手法による文書要約装置と併用しても何ら差し支えない。

【００４３】

【発明の効果】以上より明らかなように、第１の発明の文書要約装置は、文書入力部から入力された文書を文書解析部で段落,文および単語に分割し、特徴ベクトル生成部によって、単語辞書を用いて上記単語,文,段落および入力文書の特徴ベクトルを生成し、距離算出部によって、上記入力文書,段落および文の特徴ベクトル間の距離を所定の手順で算出し、文書要約生成部によって、上記各特徴ベクトル間距離に基づいて上記入力文書の要約を所定の手順で生成するので、上記特徴ベクトルを用いた入力文書の解析結果に基づいて入力文書の要約を生成できる。したがって、特定の文書形式や文脈を仮定することなく、簡単な処理によって文書における質の良い重要部を要約として抽出できる。

【００４４】すなわち、この発明によれば、入力文書中における不要な箇所を重要な箇所と誤ったり、仮定した文書形式や文脈に当て嵌まらない入力文書に対して全く非力であったりすることなく、種々のタイプの入力文書からより適切な要約を抽出できる。

【００４５】また、第２の発明の文書要約装置は、距離算出部によって、入力文書と各段落との特徴ベクトル間距離，上記入力文書と各段落毎の文との特徴ベクトル間距離，各段落と夫々の段落内の文との特徴ベクトル間距離または上記入力文書と各文との特徴ベクトル間距離を算出し、文書要約生成部によって、上記入力文書に最も近い段落，上記入力文書に最も近い各段落毎の文，各段落に最も近い夫々の段落内の文および上記入力文書に近い複数の文の少なくとも一つを入力文書の要約として選出するので、更に簡単な処理によって文書における質の良い重要部を抽出できる。

【図面の簡単な説明】

【図１】この発明の文書要約装置におけるブロック図である。

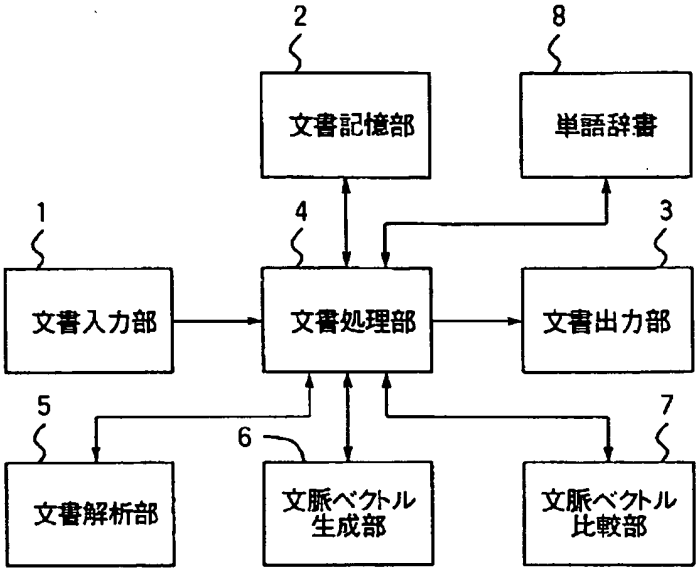
【図２】単語の文脈ベクトルが定義された単語辞書の内容の一例を示す図である。

【図３】要約作成処理動作のフローチャートである。

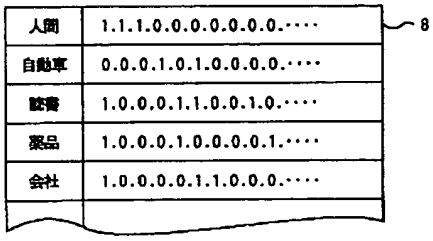
【符号の説明】

１…文書入力部、　２…文書記憶部、
３…文書出力部、　４…文書処理部、
５…文書解析部、　６…文脈ベクトル生成部、
７…文脈ベクトル比較部、　８…単語辞書。

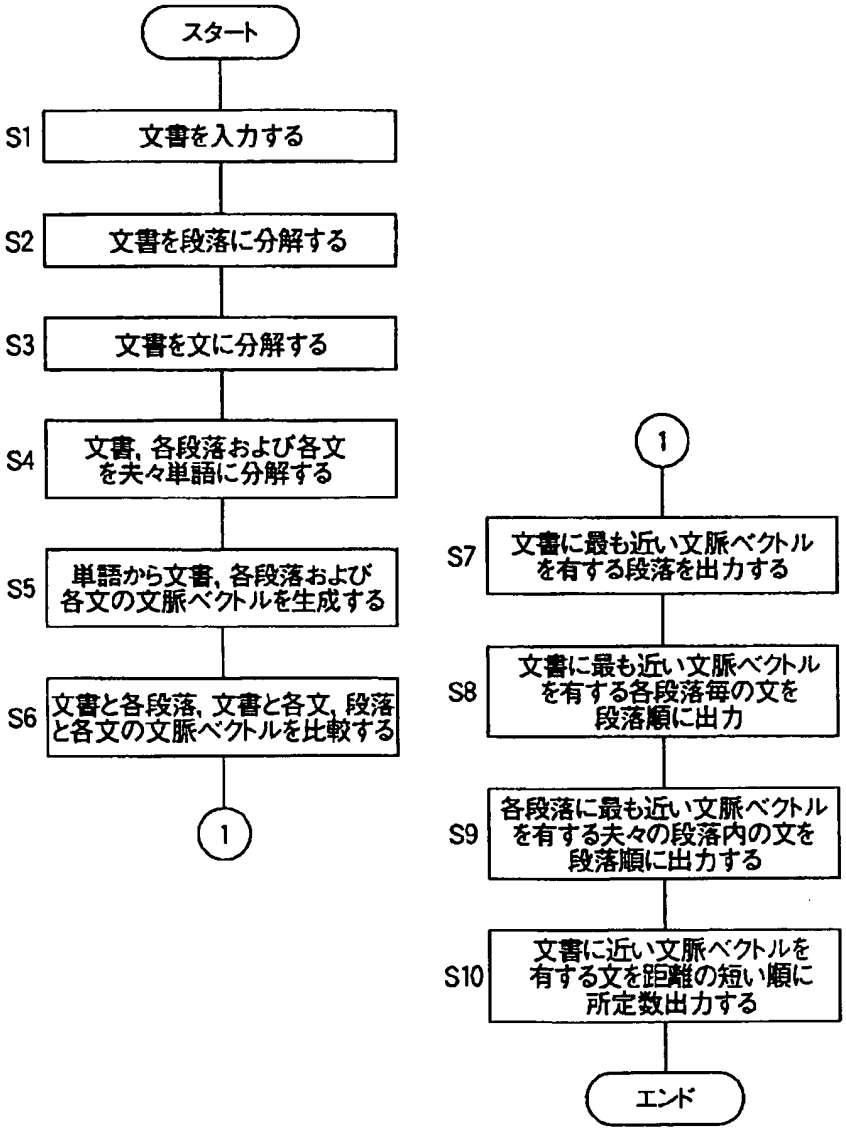
【図１】



【図２】



【図３】


スタート
S1
文書を入力する
S2
文書を段落に分解する
S3
文書を文に分解する
S4
文書，各段落および各文を夫々単語に分解する
S5
単語から文書，各段落および各文の文脈ベクトルを生成する
S6
文書と各段落，文書と各文，段落と各文の文脈ベクトルを比較する
1
1
S7
文書に最も近い文脈ベクトルを有する段落を出力する
S8
文書に最も近い文脈ベクトルを有する各段落毎の文を段落順に出力
S9
各段落に最も近い文脈ベクトルを有する夫々の段落内の文を段落順に出力する
S10
文書に近い文脈ベクトルを有する文を距離の短い順に所定数出力する
エンド